Online Supplement Materials

Interaction between FOXO1A-209 Genotype and Tea Drinking is Significantly Associated with Reduced Mortality at Advanced Ages

Yi Zeng[1,2] *, Huashuai Chen [1,3], Ting Ni [4], Rongping Ruan [5], Chao Nie[6], Xiaomin Liu[6], Lei Feng [7], Fengyu Zhang[8], Jiehua Lu[9], Jianxin Li[9], Yang Li[10], Wei Tao[11], Simon G Gregory[12], Kenneth C. Land [13], Anatoli Yashin [13], Qihua Tan [14], Ze Yang[15], Lars Bolund [6,16], Qi Ming[17, 6], Huanming Yang[6,18,19], Junxia Min[20], D. Craig Willcox [21,22], Bradley J Willcox [22], Jun Gu[11], Elizabeth Hauser[12], Xiao-Li Tian [10] *, James W. Vaupel [23].

[1] Center for the Study of Aging and Human Development, Medical School of Duke University, Durham, U.S.A.
[2] Center for Healthy Aging and Development Studies, National School of Development, Peking University, Beijing, China
[3] Department of Management, Business School of Xiangtan Universit, Xiangtan, Chinay
[4] State Key Laboratory of Genetics Engineering & MOE Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China
[5] Department of Agriculture Economics. Renmin University of China, Beijing, China
[6] Beijing Genomics Institute (BGI)-Shenzhen, Shenzhen, China
[7] Department of Psychological Medicine, National University of Singapore, Singapore
[8] Lieber Institute for Brain Development, Johns Hopkins University, Baltimore, U.S.A.
[9] Department of Sociology, Peking University, Beijing, China
[10] Department of Human Population Genetics, Institute of Molecular Medicine, Peking University, Beijing, China
[11] School of Life Sciences, Peking University, Beijing, China
[12] Duke Molecular Physiology Institute, Duke University, Durham, U.S.A.
[13] Population Research Institute, Duke University, Durham, U.S.A.
[14] Biostatistics and Biodemography, Institute of Public Health, University of Southern Denmark, Odense, Denmark
[15] National Institute of Geriatrics, Beijing Hospital, Ministry of Health of China, Beijing, China
[16] Department of Biomedicine, Aarhus University, Aarhus, Denmark
[17] Center for Genetic & Genomic Medicine, Zhejiang University School of Medicine, Hangzhou, China
[18] James D. Watson Institute of Genome Sciences, Hangzhou, China
[19] Princess Al-Jawhara Centre of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia
[20] School of Medicine, Zhejiang University, Hangzhou, China.
[21] Department of Human Welfare, Okinawa International University, Ginowan, Japan
[22] Department of Research, Kuakini Medical Center and Department of Geriatric Medicine, John A. Burns School of Medicine, University of Hawaii, Hawaii, U.S.A.
[23] Max Planck Institute for Demographic Research, Rostock, Germany

* Yi Zeng and Xiao-Li Tian are co-corresponding authors. Please address correspondence to: Yi Zeng, Duke University, Box 3003, School of Medicine, Duke University, Durham, NC 27710, U.S.A. Tel. 1-919-6607554; Fax: 1-919-668-0453; Email: zengyi@duke.edu

A technical note on how to estimate the hazard ratios of mortality risk for those

who have different combinations of the statuses of genotype and the environmental factor

An interaction between an environmental factor and a genotype is present if the association between the environmental factor and a health outcome indicator differs among individuals with different status of carrying the genotype. To understand the GxE interaction effects, regressions may be estimated separately for genotype carriers and non-carriers (dominant or recessive models), or separately for those who carry 0, 1, or 2 copies of the minor alleles (additive model), to assess the differences in effects of an environmental factor on a health outcome indicator among those who have different genotypes. If the genotype-specific regression analyses are to be carried out, however, sufficiently large sub-sample size is required for each genotype group, but such required data are likely not available in most circumstances, including our present study. Thus, we apply a simple procedure to assess the differences in effects of an environmental factor on a health outcome indicator among those who have different genotypes, without further dividing the samples. Note that this procedure was applied in previous publications in which the genotype was defined by following the dominant or recessive model and the environmental factors were binary variables in the logistic regression models. We extend here the procedure to the cases in which genotypes are defined by the additive model (or recessive or dominant model) and the environmental factors are ordered (or binary) variables, employing the Cox proportional hazards model.

In general, the Cox proportional hazards model is expressed as:

$$\log h_i(t) = \log h_0(t) + [\beta_1 G_i + \beta_2 E_i + \beta_3 G_i * \mathrm{E}_i + \sum_j \alpha_j X_{j_i}] \qquad [1]$$

where $h_i(t)$ is the hazard at time t of the i^{th} individual and $h_0(t)$ is the baseline hazard at time t; G_i represents the genotype and E_i represents environmental factor status of the i^{th} individual; $G_i x E_i$ is the interaction term of the genotype and environmental factors; X_{ji} is a vector of covariate values corresponding to the i^{th} individual. Coefficients β_1, β_2, β_3 and α_j measure the hazards of mortality risk for the corresponding variables. G_i may be a binary variable (dominant or recessive model) or an ordered variable (additive model); E_i may be a binary variable (E_i = 1 or 0 refers to exposure or not exposure to the environmental factor, such as drinking tea or not-drinking tea), or an ordered variable (E_i=0, 1, 2,…; such as E_i=0, 1, 2 refers to never, sometimes and often drinking tea).

Let HR_{GE} represent the hazard ratio of mortality risk of those with a combination of the genotype status of carrying the minor allele(s) (G) such as the FOXO1A-209 genotype and an environmental factor (E) such as tea drinking. Note that there are two limitations if one simply uses general Cox proportional hazards model expressed in Equation (1) and the coefficients β_1, β_2 and β_3 produced by the standard software to estimate the hazard ratios of mortality risk (HR_{GE}) for those individuals with combined statuses of the genotype and the environmental factor. First, we have no way to estimate the p values of HR_{GE} based on the coefficients β_1, β_2, β_3 and the other outcomes produced by the standard software. The second limitation is that it would involve an unrealistic assumption. For example, using Equation (1) to estimate HR_{GE} in

the present study would implicitly assume that the mortality risk for those who carry 2 copies (G_i=2) of the FOXO1A-209 minor allele is twice as high as that for those who carry one copy (G_i=1) of the minor allele, and assume that mortality risk for those who often drink tea (E_i=2) is twice as high as that for those who sometimes drink tea (E_i=1); these assumptions may likely not be correct in real world. To avoid these two limitations, we estimate HR_{GE} and their p values by setting up several exclusive dummy variables Z_{GE} in the regression equations. Z_{GE} represents the combinations of the genotype status (the subscript G = 0, 1 for dominant and recessive models, or G = 0, 1, 2 for additive model) and the environmental factor status (the subscript E = 0 or 1 for binary variable, or E = 0, 1, 2, … for ordered variable).

We estimated the hazard ratios of mortality risk (HR_{GE}) of the following four types of combinations of FOXO1A-209 genotypes (with recessive or additive model) and environmental factor of tea drinking (binary or ordered variable), using the Equations below, and the results are presented in Tables 3 and 4, Figures 1 and 2:

(1) FOXO1A-209 genotypes with recessive model and tea drinking binary variable:

$$\log h_i(t) = \log h_0(t) + [\gamma_{01} Z_{G_i=0,E_i=1} + \gamma_{10} Z_{G_i=1,E_i=0} + \gamma_{11} Z_{G_i=1,E_i=1} + \sum_j \alpha_j X_{ji}] \quad [2]$$

(2) FOXO1A-209 genotypes with recessive model and tea drinking ordered variable:

$$\log h_i(t) = \log h_0(t) + [\sum_{E_i=1}^{2} \gamma_{0E_i} Z_{G_i=0,E_i} + \sum_{E_i=0}^{2} \gamma_{1E_i} Z_{G_i=1,E_i} + \sum_j \alpha_j X_{ji}] \quad [3]$$

(3) FOXO1A-209 genotypes with additive model and tea drinking binary variable:

$$\log h_i(t) = \log h_0(t) + [\gamma_{0E_1} Z_{G_i=0,E_1} + \sum_{G_i=1}^{2} \sum_{E_i=0}^{1} \gamma_{G_iE_i} Z_{G_iE_i} + \sum_j \alpha_j X_{ji}] \quad [4]$$

(4) FOXO1A-209 genotypes with additive model and tea drinking ordered variable:

$$\log h_i(t) = \log h_0(t) + [\sum_{E_i=1}^{2} \gamma_{0E_i} Z_{G_i=0,E_i} + \sum_{G_i=1}^{2} \sum_{E_i=0}^{2} \gamma_{G_iE_i} Z_{G_iE_i} + \sum_j \alpha_j X_{ji}] \quad [5]$$

As shown by the results presented in Tables 3 and 4 and Figures 1 and 2, the estimates of HR_{GE} based on the regressions models expressed in Equations [2]-[5] with dummy variables Z_{GE} which represent the combinations of the statuses of genotype and the environmental factor adequately and intuitively reveal the GxE effects, although there are no explicit interaction terms in the regression equations.